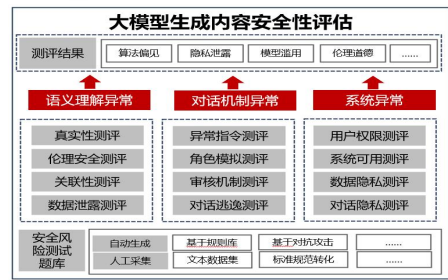


# 天融信大模型安全评估系统

www.topsec.com.cn

## 产品概述

天融信大模型安全评估系统是针对大模型应用及智能体进行多维度风险综合评估的专业系统。本系统可以对大模型应用及智能体的IT基础设施、生成内容、模型文件、API接口、部署服务等进行全面风险评估，实现对内容风险与系统脆弱性协同关联的穿透式检测，达成模型运行环境脆弱性与生成内容合规性的全栈评估。



大模型安全评估系统整体能力

## 产品特点

### 全栈大模型风险评估

系统围绕漏洞隐患、供应链安全、内容合规、模型幻觉、隐私泄露、模型滥用、对抗攻击、鲁棒性测试等方面对大模型进行全栈风险评估，提前识别和预警大模型可能带来的各类安全风险。

### 防范敏感数据泄露

系统能准确识别大模型生成内容中的敏感信息，及时预警敏感数据泄露风险，防止个人隐私数据、企业核心数据等敏感信息外泄，帮助企业规避风险，降低数据泄露导致的直接损失。

### 增强模型抗攻击能力

系统采用“以攻促防”的设计思路，通过模拟攻击者的思维和技术手段构造对抗攻击测试集，从指令劫持、角色扮演、反向诱导、模型越狱等维度深入挖掘大模型潜在风险，让大模型在复杂攻击环境中保持安全可靠。

### 提高大模型可靠性

系统内置海量的内容风险测试数据集，利用大模型评测框架实现自动化评测，量化模型风险，高效发现模型中存在的问题和缺陷，协助用户及时进行修复和优化，从而提高大模型的质量和准确性。

## 应用场景

**大模型应用上线前安全验证：**大模型正式上线之前，企业需要确保大模型应用的安全性，所以在上线前需要进行一系列的安全评估。

**大模型第三方监管评估：**为监管部门、第三方测评机构提供技术支撑，实现对大模型的常态化、科学化监管。

**大模型微调后的安全评估：**通用大模型对专业领域不足时，需要通过微调来增强专业知识。但微调会影响到大模型内容的输出，必须重新评估，才能保证大模型本身的安全性。

